Published in final edited form as:
Stat Med. 2010 May 20; 29(11): 1206–1218. doi:10.1002/sim.3862.

On the Estimation of Disease Prevalence by Latent Class Models for Screening Studies Using Two Screening Tests with Categorical Disease Status Verified in Test Positives Only

Haitao Chu[1,2,*], **Yijie Zhou**[3], **Stephen R. Cole**[4], and **Joseph G. Ibrahim**[1]

[1] Department of Biostatistics, The Univerity of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA

[2] Lineberger Comprehensive Cancer Center, The Univerity of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA

[4] Department of Epidemiology, The Univerity of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA

[3] Merck Research Laboratories, Merck & Co., Inc., Rahway, NJ 07065

Summary

To evaluate the probabilities of a disease state, ideally all subjects in a study should be diagnosed by a definitive diagnostic or gold standard test. However, since definitive diagnostic tests are often invasive and expensive, it is generally unethical to apply them to subjects whose screening tests are negative. In this article, we consider latent class models for screening studies with two imperfect binary diagnostic tests and a definitive categorical disease status measured only for those with at least one positive screening test. Specifically, we discuss a conditional independent and three homogeneous conditional dependent latent class models and assess the impact of misspecification of the dependence structure on the estimation of disease category probabilities using frequentist and Bayesian approaches. Interestingly, the three homogeneous dependent models can provide identical goodness-of-fit but substantively different estimates for a given study. However, the parametric form of the assumed dependence structure itself is not "testable" from the data, and thus the dependence structure modeling considered here can only be viewed as a sensitivity analysis concerning a more complicated non-identifiable model potentially involving heterogeneous dependence structure. Furthermore, we discuss Bayesian model averaging together with its limitations as an alternative way to partially address this particularly challenging problem. The methods are applied to two cancer screening studies, and simulations are conducted to evaluate the performance of these methods. In summary, further research is needed to reduce the impact of model misspecification on the estimation of disease prevalence in such settings.

Keywords

maximum likelihood; Bayesian inference; diagnostic test; dependence; screening; latent class models

1. Introduction

Screening for a specific disease or condition is a fundamental component of human disease control and prevention. The objective of screening is to classify asymptomatic people as

*Corresponding Author: hchu@bios.unc.edu.

likely or unlikely to have the disease or condition of interest. People who appear likely to have the disease or condition are examined further for a diagnosis, and those people who are diagnosed with the disease are treated. Therefore, screening can reduce the morbidity and mortality of the disease among people screened and can enable early treatment for diagnosed cases. Screening programs for cancer and heart diseases are well established in many countries. In many screening programs, a population with known size n is screened by two imperfect binary diagnostic tests. If the results of both diagnostic tests are negative, no further screening is undertaken. If either of the two diagnostic tests is positive, then a full evaluation of the disease using a gold standard classification is undertaken [1].

For estimating diagnostic accuracy without a gold standard, it is well known that if the conditional independence assumption is incorrectly assumed, parameter estimates may be biased [2–4]. When the disease status D is a binary random variable, Albert and Dodd [5] showed that the estimation of diagnostic accuracy and prevalence is sensitive to the choice of dependence structure for studies with multiple diagnostic tests. The dependence structure was specified using a Gaussian random effects model [6,7] and a finite mixture model [8]. They showed that it is difficult to distinguish between different dependence structures in the absence of a gold standard test in most practical situations (i.e., unless there are more than 10 tests). Albert [9] proposed methods for estimating diagnostic accuracy of multiple binary tests with an imperfect reference standard when information about the diagnostic accuracy of the imperfect test is available from external data sources. Furthermore, using the same dependence structure, Albert and Dodd [10] examined the effect of model misspecification on the estimation of test accuracy and prevalence when a binary gold standard is partially verified. They showed that for extreme biased sampling the estimation is sensitive to the choice of dependence structure. Other latent class models with a focus on diagnostic accuracy have also been considered in a single study [11,12] as well as in a meta-analysis [13]. In addition, Black and Craig [14] discussed the estimation of disease prevalence in a scenario involving two imperfect tests in the absence of a gold standard and proposed Bayesian model averaging for inference over the conditional independence and dependence models. However, those dependence models are not directly applicable in the setting that we are considering because there are only two diagnostic tests, and more importantly, if both diagnostic tests are negative, no further gold standard classification will be applied.

Let T_1, T_2, and D be the random variables denoting the two screening tests and the disease status, respectively. In this article, we consider T_1 and T_2 to be binary variables with value 1 indicating test positive and 0 indicating test negative, and D to be a categorical variable with value d = *1, 2,…, K* indicating the classes of disease. Let x_{ij}^d be the observed frequency with $D= d$, $T_1 = i$ and $T_2 = j$ ($i = 0, 1$ and $j = 0, 1$), $x_{ij}=\sum_d x_{ij}^d$ be the observed frequency with $T_1 = i$ and $T_2 = j$, and $n=\sum_i\sum_j x_{ij}$ be the total number of observations. Furthermore, let $\pi_{ij} = P(T_1 = i, T_2 = j)$, $\pi_{i+}=\sum_j \pi_{ij}, \pi_{+j}=\sum_i \pi_{ij}$, $P_d = P(D = d)$, $\pi_{ij}^d=P(D=d|T_1=i, T_2=j)$, $P_{ij}^d=P(T_1=i, T_2=j|D=d)$, $P_{i+}^d=\sum_j P_{ij}^d$, $P_{+j}^d=\sum_i P_{ij}^d$, and $P_{i|j}^d=P(T_1=i|T_2=j, D=d)$ denote the corresponding joint, marginal and conditional probabilities. In most studies, we only observe frequencies of x_{11}^d, x_{10}^d and x_{01}^d due to the nature of screening. The frequencies of x_{00}^d are usually not observed, although the margin $x_{00}=\sum_d x_{00}^d$ is observed. The data structure contains $3d+1$ observed frequencies, which in

general allows for the estimation of a maximum of 3*d* free parameters. In this paper we will not consider special cases when only less than 3*d* free parameters are identifiable, e.g., when there are zeros among the observed frequencies.

One way to write the likelihood function (ignoring constant terms) in this setting is in terms of P_d and P_{ij}^d ($d = 1, 2,\ldots, K$; $i = 0, 1$; $j = 0, 1$) with constraints of $\sum_d P_d = 1$ and $\sum_i \sum_j P_{ij}^d = 1$ as follows,

$$x_{00}\log\left(\sum_d P_{00}^d P_d\right) + \sum_d x_{11}^d \log(P_{11}^d P_d) + \sum_d x_{10}^d \log(P_{10}^d P_d) + \sum_d x_{01}^d \log(P_{01}^d P_d). \quad (1)$$

This parameterization involves a mixture likelihood in the first term and prevents a closed-form solution for the maximum likelihood estimators (MLE). It contains $4d-1$ free parameters. Without further assumptions, the parameters in equation (1) are not identifiable. However, this parameterization allows for direct specification of commonly used assumptions, usually specified through some constrains on P_{ij}^d. For example, the frequently used conditional independence assumption [15,16] assumes that the two tests T_1 and T_2 are independent conditioning on the disease status D, i.e., $T_1 \perp T_2 | D$, and the number of free parameters in equation (1) is reduced to $3d-1$ giving model identification since $P_{11}^d = P_{1+}^d P_{+1}^d$, $P_{10}^d = P_{1+}^d(1 - P_{+1}^d)$, $P_{01}^d = P_{1+}^d(1 - P_{1+}^d)$, and $P_{00}^d = (1 - P_{1+}^d)(1 - P_{+1}^d)$. For convenience, we denote the conditional independence model as the $\perp$ model with $\widehat{P}_d^{\perp}$ as the corresponding MLEs. Under the homogeneous dependence assumptions (i.e., the α, θ, and ρ models that will be discussed in Sections 2 and 3), the number of free parameters in equation (1) is reduced to 3*d* and the models become saturated and equivalent to the alternative parameterization below [17].

An alternative parameterization of the log-likelihood function can be written in terms of π_{ij} ($i = 0, 1$; $j = 0, 1$) and π_{ij}^d ($i+j>0$ and $d=1, 2, \ldots, K$) with constraint of $\sum_i \sum_j \pi_{ij} = 1$ as follows,

$$\sum_i \sum_j x_{ij}\log(\pi_{ij}) + \sum_d x_{11}^d \log(\pi_{11}^d) + \sum_d x_{10}^d \log(\pi_{10}^d) + \sum_d x_{01}^d \log(\pi_{01}^d). \quad (2)$$

This representation relates to previous work in other settings [18–20]. This model is a saturated model with *3d* parameters. The maximum likelihood equations are tractable and yield MLEs in closed-form. Omitting the algebra, we obtain, $\hat{\pi}_{ij} = x_{ij}/n$ ($i, j = 0, 1$), and $\widehat{\pi}_{ij}^d = x_{ij}^d / x_{ij}$ if $i + j>0$. The existence of closed-form solutions for this alternative parameterization allows for closed- form solutions for the α and θ homogeneous conditional dependent models [17], which will be briefly discussed in detail in Sections 2.1 and 2.2. Furthermore, this saturated alternative parameterization also suggests that the probability of having disease class *d* for those with both tests negative (i.e., π_{00}^d), and further the overall probability of having disease class *d* (i.e. $P_d = \sum_i \sum_j \pi_{ij}\pi_{ij}^d$) are not identifiable without some additional "non-testable" assumptions. Thus, the dependence structure modeling considered

in this paper itself is not "testable", and can only be viewed as a sensitivity analysis for the estimation of disease prevalence. Similar to generalized linear models with non-ignorable missing data mechanism [21], the type of sensitivity analyses play an important role in the estimation and inference in this problem.

In similar settings when the gold standard was only measured on those who screened positive, Cheng et al. [22] and Pepe and Alonzo [23] have examined the potential overwhelming impact of the correlation between the two screening tests on the estimation of absolute test accuracy parameters. Both suggest using relative test accuracy for comparing disease screening tests. However, to our knowledge, no one has assessed the impact of the misspecification of conditional dependence structures, which can be specified by a homogeneous dependence parameter for two diagnostic tests, on the estimation of disease class probabilities in such screen-positive ascertained studies.

In this article, we empirically assess the impact of misspecification of the conditional dependence structure on the estimation of disease class probabilities through two case studies and simulations. Specifically, in Sections 2.1 and 2.2, we define the MLEs for the homogeneous dependent α and θ models, and in Section 2.3 we propose the homogeneous correlation coefficient conditional dependent ρ model. Bayesian approaches, which incorporate prior beliefs about dependence, are developed for the three models in Section 3 as an alternative to the maximum likelihood methods. Furthermore, we discuss Bayesian model averaging in Section 3.4 as an alternative way to address the challenging estimation problem since the three homogeneous dependent models can provide the same goodness-of-fit for the data but substantively different estimates [17] and the dependence structure itself is not "testable". In Section 4, we compare the results for the two case studies using both the maximum likelihood methods and the Bayesian approaches. The two case studies were reanalyzed recently by Böhning and Patilea [1] using a capture-recapture approach under the α and θ model assumptions. Our focus here is to compare the estimates under the α, θ and ρ model assumptions using both the maximum likelihood methods and Bayesian approaches. A simulation study is conducted in Section 5 and a brief discussion is presented in Section 6.

2. Maximum Likelihood Estimators for Models with Homogeneous Dependence

In Section 2.1 and 2.2, we will briefly introduce the homogeneous conditional dependence α and θ models, recently proposed by Böhning and Patilea [1] using a capture-recapture approach. Using the alternative parameterization of the maximum likelihood as presented in equation (2), Chu and Nie [17] presented closed-form maximum likelihood solutions under the α and θ model assumptions.

2.1 Homogeneous conditional dependence: the *α* model

Under this model, the association of the two tests T_1 and T_2 conditional on the disease status D as measured by the odds ratio is assumed to be homogeneous over all disease categories, i.e., $\alpha_d = \frac{P_{11}^d P_{00}^d}{P_{01}^d P_{10}^d}$ ($d = 1, 2, \ldots, K$) is assumed to be homogenous. By Bayes' theorem, we obtain $\alpha_d = \frac{\pi_{11}\pi_{00}}{\pi_{01}\pi_{10}} \times \frac{\pi_{11}^d}{\pi_{01}^d \pi_{10}^d} \pi_{00}^d$. With $\sum_d \pi_{00}^d = 1$ and simple algebra, we obtain the solution of $\alpha = \frac{\pi_{11}\pi_{00}}{\pi_{01}\pi_{10}} \times \left[\sum_d \frac{\pi_{01}^d \pi_{10}^d}{\pi_{11}^d} \right]^{-1}$ under this homogeneity assumption. Thus, by plugging in the

closed-form solutions of MLEs of π ($i, j = 0, 1$) and π_{ij}^d $(i+j>0)$ from equation (2), the closed-form MLEs of α and P_d^α are

$$\widehat{\alpha}=n\times\frac{x_{11}^+x_{00}^+}{x_{01}^+x_{10}^+}\times\left(\sum_d\frac{x_{01}^dx_{10}^d}{x_{11}^d}\right)^{-1},\ \widehat{P}_d^\alpha=\frac{1}{n}\left[x_{11}^d+x_{10}^d+x_{01}^d+x_{00}\frac{x_{01}^dx_{10}^d}{x_{11}^d}\left(\sum_d\frac{x_{01}^dx_{10}^d}{x_{11}^d}\right)^{-1}\right], \tag{3}$$

where the superscript α indicates the homogeneous odds ratio assumption.

2.2 Homogeneous conditional dependence: the *θ* model

Under this model, ratio of conditional (conditional on test T_2 being positive) and unconditional probabilities of test T_1 being positive is assumed to be homogeneous over all disease categories, i.e., $\theta_d=\frac{P_{1|1}^d}{P_{1+}^d}$ ($d = 1, 2, \ldots, K$) is assumed to be homogenous. By Bayes' theorem, we obtain $\theta_d=\frac{\pi_{11}\pi_{11}^d}{(\pi_{01}\pi_{01}^d+\pi_{11}\pi_{11}^d)(\pi_{10}\pi_{10}^d+\pi_{11}\pi_{11}^d)}\times P_d$. With $\sum_d P_d=1$ and simple algebra, we obtain the solution of $\theta=\left[\sum_d\frac{1}{\pi_{11}\pi_{11}^d}(\pi_{10}\pi_{10}^d+\pi_{11}\pi_{11}^d)(\pi_{01}\pi_{01}^d+\pi_{11}\pi_{11}^d)\right]^{-1}$. Thus, the closed- form MLEs of θ and P_d^θ are

$$\widehat{\theta}=n\left[\sum_d\frac{1}{x_{11}^d}(x_{10}^d+x_{11}^d)(x_{01}^d+x_{11}^d)\right]^{-1},\ \widehat{P}_d^\theta=\frac{1}{n}\left[\frac{x_{1+}^dx_{+1}^d}{x_{11}^d}\left(\sum_d\frac{x_{1+}^dx_{+1}^d}{x_{11}^d}\right)^{-1}\right], \tag{4}$$

where the superscript θ indicates the homogeneous relative risk assumption.

2.3 Homogeneous conditional dependence: the ρ model

In this section, we propose an alternative homogeneous conditional dependence model, the ρ model. Under this model, the correlation of the two tests T_1 and T_2 is assumed to be homogeneous over all disease categories, i.e., ρ_d($d = 1, 2, \ldots, K$) is assumed to be homogenous ρ. Let $\delta_d=\rho_d\sqrt{P_{1+}^dP_{+1}^d(1-P_{1+}^d)(1-P_{+1}^d)}$ be the covariance between two tests in the d^{th} disease group, then we have $P_{11}^d=P_{1+}^dP_{+1}^d+\delta_d$, $P_{10}^d=P_{1+}^d(1-P_{1+}^d)-\delta_d$, $P_{01}^d=P_{+1}^d(1-P_{1+}^d)-\delta_d$, and $P_{00}^d=(1-P_{1+}^d)(1-P_{+1}^d)+\delta_d$. The bounded range of correlations is determined by the marginal probability of testing positive P_{1+}^d and P_{+1}^d. Specifically, the correlation coefficients ρ satisfies

$$\max_d\left\{-\sqrt{\frac{P_{+1}^dP_{1+}^d}{(1-P_{+1}^d)(1-P_{+1}^d)}},-\sqrt{\frac{(1-P_{+1}^d)(1-P_{+1}^d)}{P_{+1}^dP_{1+}^d}}\right\}\le\rho\le\min_d\left\{\sqrt{\frac{(1-P_{+1}^d)P_{1+}^d}{P_{+1}^d(1-P_{1+}^d)}},\sqrt{\frac{P_{+1}^d(1-P_{1+}^d)}{(1-P_{+1}^d)P_{1+}^d}}\right\}. \tag{5}$$

Let the MLEs of ρ and P_d be denoted as $\hat{\rho}$ and $\widehat{P}_d^{\rho}$, where the superscript ρ indicates the homogeneous correlation coefficient assumption. They do not have a closed-form solution under this model.

All three models assume a homogeneous dependence structure. This is a rather strong assumption. However, because all three homogeneous dependent models are already saturated, heterogeneous dependent models are not identifiable without additional constraints on test accuracy parameters (e.g., assuming the test accuracy parameters are the same for the two diagnostic tests, which is a much stronger assumption in general). Furthermore, the three homogeneous dependent models can provide the same goodness-of-fit for the data but substantively different estimates [17], a natural way addressing this problem might be through the frequentist model average estimators [24]. Let w_α, w_θ and w_ρ with constraint of $w_\alpha + w_\theta + w_\rho = 1$ be the corresponding weights for the α, the θ, and the ρ models, the weighted model average estimator can be defined as $P_d^w = w_\alpha P_d^\alpha + w_\theta P_d^\theta + w_\rho P_d^\rho$. However, the MLEs $\widehat{P}_d^{\alpha}$, $\widehat{P}_d^{\theta}$ and $\widehat{P}_d^{\rho}$ are usually correlated since they are based on the same data. Due to the technical difficulty of computing the variance-covariance matrix between ($\widehat{P}_d^{\alpha}$, $\widehat{P}_d^{\theta}$) and $\widehat{P}_d^{\rho}$ for the computation of the standard error of $\widehat{P}_d^{w}$, we do not consider the frequentist model average estimator in this article. We will consider the Bayesian model averaging counterpart in Section 3.4.

In practice, it is often of interest to test the difference between the estimated probabilities of disease states using different dependence assumptions (i.e., the α, θ or ρ model). Since the closed-form maximum likelihood solutions for the α and θ models are based on the likelihood function as presented in equation (2), a Wald-type test comparing $\widehat{p}_d^{\alpha} - \widehat{p}_d^{\theta}$ is directly available with the standard error $se\left(\widehat{p}_d^{\alpha} - \widehat{p}_d^{\theta}\right)$ obtained by the delta method. Due to the technical difficulty of computing the variance-covariance matrix between ($\widehat{P}_d^{\alpha}$, $\widehat{P}_d^{\theta}$) and $\widehat{P}_d^{\rho}$, comparing the MLEs of ($\widehat{P}_d^{\alpha}$, $\widehat{P}_d^{\theta}$) with $\widehat{P}_d^{\rho}$ is not straightforward. In practice, bootstrapping methods can be used as an alternative way to compute the corresponding p-values and 95% confidence intervals [25].

We developed a SAS macro (SAS Institute, Cary, NC) to implement the models discussed above parameterized both in terms of P_d and P_{ij}^d as in equation (1) for the homogeneous ρ model, and in terms of π_{ij} and π_{ij}^d as in equation (2) for the homogenous α and θ models. To describe disease class prevalence and to implement the constraints of $0 < P_d < 1$ and $\sum_d P_d = 1$, we used the linear generalized logit model [26] which is widely applied in categorical data analysis. This model has an inverse link function defined as $P_d = \frac{\exp(\beta_d)}{\sum_d \exp(\beta_d)}$ (d = 1, 2, …, K) with $\beta_K = 0$. We used the delta method to compute the standard error of functions of MLEs and their confidence intervals based on normal approximation. The two parameterizations (i.e., in terms of P_d, P_{ij}^d and in terms of π_{ij}, π_{ij}^d) provide exactly the same results.

3. Bayesian Estimation for Models with Homogeneous Dependence

In this Section, we discuss the Bayesian approaches [27,28]. Because the Bayesian approach and the frequentist approach use different frameworks, they can be considered

complementary. When relatively large studies are combined with weak prior distributions, inferences obtained by Bayesian and frequentist methods generally agree. However, the Bayesian framework is particularly attractive when suitable prior distributions can be constructed to incorporate known constraints and subject-matter knowledge on model parameters [29]. The Bayesian framework allows direct construction of 100(1−α)% equal tail and highest probability density (HPD) credible intervals of general functions of the estimated parameters without having to rely on asymptotic approximations. Furthermore, the Bayesian framework provides direct implementation of model averaging [30], which provides a natural way to address the problem of selecting a model from several competing models that give equal goodness-of-fit but potentially different inferences for a particular study.

3.1 Homogeneous conditional dependence: the *α* model

To implement the constrains of $\sum_i \sum_j P_{ij}^d = 1$ and $\alpha_d = \frac{P_{11}^d P_{00}^d}{P_{01}^d P_{10}^d} = \alpha$ under the α model, we re-parameterize $P_{11}^d, P_{01}^d, P_{10}^d$ and P_{00}^d as follows,

$$P_{11}^d = \frac{\alpha \exp(a_d + b_d)}{1 + \exp(a_d) + \exp(b_d) + \alpha \exp(a_d + b_d)}, \; P_{10}^d = \frac{\exp(a_d)}{1 + \exp(a_d) + \exp(b_d) + \alpha \exp(a_d + b_d)}$$
$$P_{01}^d = \frac{\exp(b_d)}{1 + \exp(a_d) + \exp(b_d) + \alpha \exp(a_d + b_d)}, \; P_{00}^d = \frac{1}{1 + \exp(a_d) + \exp(b_d) + \alpha \exp(a_d + b_d)}. \quad (6)$$

Let $f(\alpha, a_d, b_d, P_d)$ be the prior joint distribution of (α, a_d, b_d, P_d) $(d = 1, 2, \ldots, K)$, the joint posterior distribution given the observed frequencies is proportional to

$$(P_d)^{x_{11}^d + x_{10}^d + x_{01}^d} (P_{11}^d)^{x_{11}^d} (P_{10}^d)^{x_{10}^d} (P_{01}^d)^{x_{01}^d} \left(\sum_d P_d P_{00}^d \right)^{x_{00}} f(\alpha, a_d, b_d, P_d) = \frac{(P_d)^{x_{11}^d + x_{10}^d + x_{01}^d} \alpha^{x_{11}^d} \exp\left[a_d (x_{11}^d + x_{10}^d) + b_d (x_{11}^d + x_{01}^d) \right]}{\left[1 + \exp(a_d) + \exp(b_d) + \alpha \exp(a_d + b_d) \right]^{x_{11}^d + x_{10}^d + x_{01}^d}}$$
$$\times \left(\sum_d \frac{P_d}{1 + \exp(a_d) + \exp(b_d) + \alpha \exp(a_d + b_d)} \right)^{x_{00}}$$
$$\times f(\alpha, a_d, b_d, P_d) \times I(\alpha > 0). \quad (7)$$

3.2 Homogeneous conditional dependence: the *θ* model

To implement the constrains of $\sum_i \sum_j P_{ij}^d = 1$ and $\theta_d = P_{1|1}^d / P_{1+}^d = \theta$ under the θ model, we re-parameterize $P_{11}^d, P_{01}^d, P_{10}^d$ and P_{00}^d as follows,

$$P_{11}^d = \theta P_{1+}^d P_{+1}^d, \; P_{10}^d = P_{1+}^d (1 - \theta P_{+1}^d), \; P_{01}^d = P_{+1}^d (1 - \theta P_{1+}^d), \; P_{00}^d = 1 - P_{1+}^d - P_{+1}^d + \theta P_{1+}^d P_{+1}^d.$$

Let $f(\theta, P_{1+}^d, P_{+1}^d, P_d)$ be the prior joint distribution of $(\theta, P_{1+}^d, P_{+1}^d, P_d)$ $(d = 1, 2, \ldots, K)$, the joint posterior distribution given the observed frequencies is proportional to

$$
\begin{aligned}
&(P_d)^{x_{11}^d+x_{10}^d+x_{01}^d}(P_{11}^d)^{x_{11}^d}(P_{10}^d)^{x_{10}^d}(P_{01}^d)^{x_{01}^d}\left(\sum_d P_d P_{00}^d\right)^{x_{00}} f(\theta,\\
&P_{1+}^d, P_{+1}^d, P_d)=(P_d)^{x_{11}^d+x_{10}^d+x_{01}^d}(P_{1+}^d)^{x_{11}^d+x_{10}^d}(P_{+1}^d)^{x_{11}^d+x_{01}^d}\theta^{x_{11}^d}(1-\theta P_{+1}^d)^{x_{10}^d}(1-\theta P_{1+}^d)^{x_{01}^d}\\
&\times\left[\sum_d P_d(1-P_{1+}^d-P_{+1}^d+\theta P_{1+}^d P_{+1}^d)\right]^{x_{00}}\\
&\times f(\theta,\\
&P_{1+}^d, P_{+1}^d, P_d)\times I(\theta>0)\times I(1\\
&-\theta P_{+1}^d>0)\\
&\times I(1\\
&-\theta P_{1+}^d>0)\\
&\times I(1\\
&-P_{1+}^d-P_{+1}^d+\theta P_{1+}^d P_{+1}^d>0).
\end{aligned}
\tag{8}
$$

The feasible range of θ is determined by the marginal probability of testing positive P_{1+}^d and P_{+1}^d and is implemented through the addition of the four indicator functions I(·) in equation (8).

3.3 Homogeneous conditional dependence: the ρ model

To implement the constrains of $\sum_i\sum_j P_{ij}^d=1$ and $\rho_d=\rho$ ($d = 1, 2, \ldots, K$), we re-parameterize $P_{11}^d, P_{01}^d, P_{10}^d$ and P_{00}^d as in equation (6). Let $f(\rho, P_{1+}^d, P_{+1}^d, P_d)$ be the prior joint distribution of $(\theta, P_{1+}^d, P_{+1}^d, P_d)$ ($d = 1, 2, \ldots, K$) and $\delta_d=\rho_d\sqrt{P_{1+}^d P_{+1}^d(1-P_{1+}^d)(1-P_{+1}^d)}$ be the covariance, the joint posterior distribution given the observed frequencies is proportion to

$$
\begin{aligned}
&(P_d)^{x_{11}^d+x_{10}^d+x_{01}^d}(P_{11}^d)^{x_{11}^d}(P_{10}^d)^{x_{10}^d}(P_{01}^d)^{x_{01}^d}\left(\sum_d P_d P_{00}^d\right)^{x_{00}^+} f(\rho,\\
&P_{1+}^d, P_{+1}^d, P_d)=(P_d)^{x_{11}^d+x_{10}^d+x_{01}^d}(P_{1+}^d P_{+1}^d+\delta_d)^{x_{11}^d}(P_{1+}^d - P_{1+}^d P_{+1}^d - \delta_d)^{x_{10}^d}(P_{+1}^d - P_{+1}^d P_{1+}^d - \delta_d)^{x_{01}^d}\\
&\times\left\{\sum_d P_d\left[(1-P_{1+}^d)(1-P_{+1}^d)+\delta_d\right]\right\}^{x_{00}}\\
&\times f(\rho,\\
&P_{1+}^d, P_{+1}^d, P_d)\times I(P_{1+}^d P_{+1}^d\\
&+\delta_d>0)\times I(P_{1+}^d\\
&-P_{1+}^d P_{+1}^d\\
&-\delta_d>0)\\
&\times I(P_{+1}^d\\
&-P_{+1}^d P_{1+}^d\\
&-\delta_d>0)\\
&\times I(1\\
&-P_{1+}^d - P_{+1}^d + P_{1+}^d P_{+1}^d + \delta_d>0).
\end{aligned}
\tag{9}
$$

The feasible range of correlation determined by the marginal probability of test positive P_{1+}^d and P_{+1}^d as in equation (5) is implemented through the addition of the four indicator functions I(·) in equation (9).

3.4 Homogeneous conditional dependence: Bayesian model averaging (BMA)

The homogeneous dependence models are saturated. Therefore, they provide the same goodness-of-fit for the data, but can provide substantively different estimates. Bayesian model averaging (BMA) provides a natural way to address this problem [30]. The posterior distribution of the quantity of interest P_d given data is

$$
pr(P_d|Data)=\sum_{k=1}^{K} pr(P_d|M_k, Data)\,pr(M_k|Data),
\tag{10}
$$

where $M_1 \ldots, M_K$ are the models considered, and the posterior probability for model M_k is given by $pr(M_k|Data)=\frac{pr(Data|M_k)pr(M_k)}{\sum_{i=1}^{K} pr(Data|M_i)pr(M_i)}$, where $pr(Data|M_k)=\int pr(Data|\vartheta_k, M_k)pr(\vartheta_k|M_k)d\vartheta_k$ is the integrated likelihood of model M_k, and ϑ_k is the vector of parameters of model M_k, $pr(\vartheta_k|M_k)$ is the prior density of ϑ_k under model M_k, $pr(Data|\vartheta_k, M_k)$ is the likelihood, and $pr(M_k)$ is the prior probability that M_k is the true model (assuming one of the models considered is true). In this paper, we assume equal prior probabilities for the α, θ and ρ models, i.e., $pr(M_k)=1/3$ for k=1,2,3.

In the Bayesian models discussed above, computation was done using Markov chain Monte Carlo (MCMC) [31] in WinBUGS [32] and BRUGs in R (http://www.r-project.org). Burn-in consisted of 50,000 iterations; 50,000 subsequent iterations were used for posterior summaries. Convergence of Markov chains was assessed using the Gelman and Rubin

convergence statistic [33,34]. To describe disease class prevalence and to implement the constrain of $0 < P < 1$ and $\sum_d P_d = 1$, we use the linear generalized logit model which has inverse link function defined as $P_d = \frac{\exp(\beta_d)}{\sum_d \exp(\beta_d)}$ ($d = 1, 2, \ldots, K$) with $\beta_K = 0$ [26]. We selected proper but diffuse prior distributions for the hyperparameters [35]. Specifically, the hyper-priors for the parameters were assumed to be as follows: 1) Vague priors of $N(0, 10^3)$ were assumed for β_d s ($d = 1, 2, \ldots, K-1$) in the generalized logit transformed probabilities of disease classes P_d s; 2) Uniform prior of $[-1.0, 1.0]$ was assumed for correlation coefficient ρ; 3) Vague priors of $N(0, 10^3)$ were assumed for α and θ on the log scale to ensure $\alpha > 0$ and $\theta > 0$; and 4) Vague priors of $N(0, 10^3)$ were assumed for a_d s and b_d s in the α model to directly implement the homogeneous odds ratios assumption, and for P^d_{1+}s and P^d_{+1}s in the logit scale for the θ and ρ models.

4. Two Case Studies

For the purpose of comparing the performance of different models, we reanalyzed the data from two screening studies, in which the disease status has been evaluated only for those who tested positive for at least one of the two tests. The first study consists of data from the Health Insurance Plan Study for screening breast cancer in New York [36]. The study was carried out by the Health Insurance Plan, a prepaid comprehensive medical care plan with 750,000 subscribers enrolled in 31 medical groups. Periodic screening for breast cancer using mammography as well as clinical physical examination was performed for women aged 40 to 64 years who were chosen at random. In this study, 307 out of 20,211 women, who were test positive by either physical examination or mammography, underwent biopsy for the classification of two disease states: no cancer ($d = 1$) or cancer ($d = 2$). The second study is the multicenter study comparing cervicography with the standard pap smear cytology test for detecting cervical cancer between November 1991 and December 1992 [37]. In this study, 228 out of 5,192 women, who were test positive by either cervicography or the standard pap smear cytology test, underwent biopsy for the classification of three disease states: not present ($d = 1$), low grade (condyloma) ($d = 2$) and high grade (invasive cancer) ($d = 3$). Table 1 presents the observed frequencies in the two screening studies.

Table 2 presents the estimates of the conditional dependence parameters (i.e., α, θ and ρ) when using both the maximum likelihood method and the Bayesian method. We use the triple of percentiles, $_{2.5}50_{97.5}$, to display a parameter estimate (or posterior median) with its 95% confidence (or credible) interval, as suggested by Louis and Zeger [38]. In summary, both approaches suggest statistically significant dependence when using all three models for the two studies. Tables 3 and 4 present the estimated probabilities of the disease classes under the three homogenous dependence models as well as under the independence model, when using the maximum likelihood method and the Bayesian method, respectively. The twice negative likelihood is presented in Table 3 for comparing the goodness-of-fit of the independent ⊥ model, and the homogeneous dependent α, θ, and ρ models, which demonstrate that the α, θ, and ρ models give exactly the same goodness-of-fit for both studies. In addition, the BMA estimates across the three conditional dependence models are presented in Table 4. In summary, the estimates were consistent between the maximum likelihood and Bayesian approaches except for the probabilities of not present and low grade cervical cancer in the multicenter study detecting cervical cancer using the ρ model, potentially due to the constrains implemented in the Markov chain Monte Carlo samplings. Specifically, the estimated probability of low grade cervical cancer is estimated to

be $_{54}255_{883}$ per 1000 women using the Bayesian approach, but only $_{0}115_{308}$ per 1000 women using the maximum likelihood method.

As an interesting observation, we found that the difference between the estimated probabilities of disease states using different dependence assumptions (i.e., the α, θ or ρ model) can be statistically significant and practically meaningful. For example, in the Health Insurance Plan Study for breast cancer screening in New York, the estimated probability of having breast cancer using the maximum likelihood method is $_{3}48_{93}$ per one thousand women assuming the α model, while the estimate is $_{28}75_{122}$ per one thousand women assuming the θ model, and $_{2}7_{11}$ per one thousand women assuming the ρ model. The difference between the estimated probabilities of having breast cancer assuming α and θ models is $_{14}27_{40}$ per one thousand women with a p-value less than 0.001 by a Wald-type test. The non-overlapping 95% confidence intervals between the estimated probabilities assuming the ρ model and the α (or θ) model suggests a statistically significant difference at least at the 5% significant level. In addition, using the maximum likelihood approach, the estimated probability of having invasive high grade cervical cancer is $_{28}61_{94}$ per one thousand in the multicenter study for detecting cervical cancer assuming the θ model, which is about eight times higher than the estimate of $_{5}8_{11}$ per one thousand women assuming the ρ model, and the 95% confidence intervals do not overlap. The Bayesian approaches gave similar inferences to the frequentist approaches. This substantial difference in estimated probability high grade cervical cancer can have an impact on cancer surveillance and prevention. Unfortunately, the data does not contain any information to differentiate those dependent models since they all give the same goodness-of-fit. Thus, without some sensible assumptions, the disease prevalence may not be estimable from the data set, even with Bayesian model averaging, particularly if proposed models in BMA do not contain the correct model (which is arguably true in practice given that an infinite large number of models exist and potentially many can give same goodness of fit).

5. Simulation Studies

To further study how the disease status probability estimates vary with the dependent model assumption and to evaluate the impact of misspecification of different dependent models on the estimation of the probabilities of disease classes, we performed four sets of simulations assuming the independent model, the α, θ, and ρ dependent models, respectively. For ease of presentation and interpretation, we considered two disease strata. The simulation parameters are: the probabilities of disease classes P_d = (0.8,0.2), the marginal conditional probabilities of test T_1 being positive $P^d_{1+}=(0.10, 0.25)$, and the marginal conditional probabilities of test T_2 being positive $P^d_{+1}=(0.05, 0.30)$. In the α and θ models, we used two values of α (or θ) = 1.25 and 3.0. In the ρ model, we used two values of ρ = 0.2 and 0.6. The sample sizes considered were n = 5000 and 25000. For each combination of α (or θ, or ρ) and n values, we generated 2,000 replicates. For each replicate, we computed the estimators under the independent model, the dependent α, θ, and ρ models, using both the maximum likelihood and Bayesian approaches. In addition, the BMA estimators across the three dependent models were computed. We used the true values of P_d, P^d_{1+} and P^d_{+1}, α = 1, θ =1, and ρ = 0 as the starting values in the maximum likelihood optimization procedures and the Bayesian Markov chain Monte Carlo sampling procedures.

Table 5 presents the means of the estimated disease prevalence across 2,000 replicates, using both the maximum likelihood and Bayesian approaches. For the Bayesian models, posterior medians were used as estimates for disease prevalence for a single replicate. If the true underlying model is the conditional independence model, fitting the α, θ and ρ dependence models will provide unbiased estimates for the disease prevalence. However, if the

underlying model is one of the three dependence models, assuming independence provides biased estimates for disease prevalence. In addition, if the underlying model is a dependence model, assuming an incorrect dependence structure leads to biased estimates for disease prevalence. One interesting observation is that Bayesian model averaging (BMA) estimates tend to be less biased than the estimates under a misspecified dependence model. Furthermore when the underlying model is the α dependent model, the BMA estimates lead to nearly unbiased estimates. For all the scenarios, the maximum likelihood and Bayesian approaches provide similar estimates.

Table 6 presents the average length of the 95% confidence/credible intervals or the precision of the disease prevalence estimates across 2,000 replicates when using the maximum likelihood and Bayesian approaches. We found that if the true underlying model is the conditional independence model, assuming the α and θ dependence models leads to intervals that are too wide. For example, the 95% confidence/credible interval length using the θ dependence model is about twice than that using the true independence model. This suggests a substantive efficiency loss when conservatively assuming the α and θ dependence models. However, if the ρ dependence model is assumed, the average interval lengths are only slightly inflated. On the other hand, if the underlying model structure is one of the three dependence models, assuming independence leads to intervals that are too narrow (and biased). In addition, if the underlying model is the θ model, incorrectly assuming the α and ρ dependence models also leads to underestimation of the interval length. Furthermore, the results are highly concordant between the maximum likelihood and Bayesian approaches. Note that the average interval lengths of the BMA estimates are generally larger than those under any dependence model alone, regardless of whether the dependence model is correctly or incorrectly specified. This is due to the fact that the BMA estimates incorporate the additional uncertainty from model specification.

Table 7 presents the coverage performance of the 95% confidence/credible intervals of the disease prevalence across 2,000 replicates using both the maximum likelihood and Bayesian approaches. The coverage upon misspecification using dependent models is still around 95% if the true underlying model is the conditional independence model, possibly due to the negligible bias and wider confidence/credible intervals upon such misspecification, as suggested in Tables 5 and 6. However, if the underlying model structure is one of the three dependent models, the coverage upon misspecification decreases as the degree of dependence increases and as the sample size increases. In addition, the results suggest that if the underlying model is the ρ model, decent coverage tends to be difficult when the model is misspecified. In general, the Bayesian 95% credible intervals show slightly better coverage compared with the maximum likelihood 95% confidence intervals. More importantly, the coverage of the BMA intervals generally exceeds 90%, which has much better performance than the intervals from any single misspecified model. One reason for the better coverage using BMA is that such intervals are generally wider than those under a single model alone, and the true underlying model is included in the model averaging.

6. Discussion

For screening studies where a categorical disease status is verified only if at least one out of the two binary screening tests being positive, we investigated three homogeneous dependence models (i.e., the α, θ, and ρ models) with two case studies and four sets of simulation studies, in which the ρ model is proposed in this paper. If the true underlying model is the conditional independence ⊥ model, assuming the α and θ dependence models leads to intervals that are too wide (i.e., the 95% confidence/credible interval length of the θ model can be as twice as that of the ⊥ model), while the ρ dependence model only slightly inflated the average interval lengths. By two real data analyses and simulation studies, we

demonstrated that the three homogeneous dependence models can provide substantively different estimates for a study although with the same goodness-of-fit. We discussed both the frequentist and Bayesian approaches, and evaluated the impact of model misspecification on the estimation of disease class probabilities. Furthermore, we discussed Bayesian model averaging as an alternative way to partially address this particularly challenging estimation problem. Although we focused on the inference of disease class probabilities in this article, the same conclusion applies to the inference of the cell probabilities for two negative tests, i.e., π_{00}^{d}, and the unknown cell frequency x_{00}^{d}. We did not discuss the impact of misspecification of dependence structures on the estimation of test accuracies because it has been well studied from frequentist perspective [5,10,39]. It might be of interest to compare the performance of frequentist and Bayesian approaches on the estimation of test accuracy parameters under different settings such as low, moderate and high sensitivities and specificities.

The results imply that large differences in the estimated disease class probabilities may occur when assuming different dependence models, which can have a substantial impact on disease surveillance and prevention. Other more robust statistical methods, e.g. the generalized estimation equations [40,41], may be used to reduce the impact from misspecification of the dependence structure in this setting. We do not intend to suggest that these homogeneous dependence models are useless in practice because we cannot statistically differentiate between them based on the data alone. Caution against using these models due to the possible misspecification should be balanced with the need to estimate disease status probabilities. Furthermore, we realize that there are many more potential dependence structures than what we have considered, e.g., one could argue that the tests are dependent only for the cases but are independent for the controls [42]. Depending on the problem in hand, some assumptions may be justifiable and preferable. In addition, note that the indistinguishable characteristic of these models is based on goodness-of-fit statistics alone. We can always use additional information such as expert opinion, historic information on sensitivities and specificities of the two binary diagnostic tests, and/or the range of dependence parameters to assist our choice of selecting a homogeneous dependence model. For the Bayesian approach, the additional information can be formulated as informative priors to improve the posterior inference. However, how to solicit and formulate informative priors in this case deserves thorough investigation and is beyond our current scope.

A potential strategy to justify the homogeneous assumptions of the α, θ and ρ is to incorporate a design element into the screening study that allows the selection of homogeneity models. This strategy could be randomly selecting a subset of both test negatives for ascertainment by a gold standard. However, in cases when a gold standard test is invasive and/or expensive, it is generally considered unethical to apply it to subjects whose screening tests are negative. In this case, if historical data or additional sample from a set of confirmed cases and controls in a similar population is available for determining test accuracy parameters, one can use the data to guide the selection of homogeneity dependence models.

In cases when there is no scientific justification to prefer a particular dependence model over the others, we suggest to treat those dependence models (including the three homogeneous dependence models that we have considered) as sensitivity analyses, and investigate how the dependence structure will impact the estimation of probabilities of disease classes. If there is a clinically significant difference, caution should be taken with any statistical inference. As a last choice, if the dependence structure assumption cannot be reasonably determined, Bayesian model averaging (BMA) may be preferable to any single model, but there is a

heavy price to pay for the BMA: 1) the computations become more complex and 2) the credible intervals get much larger in some cases (to reflect the added uncertainty).

Assuming that models used in the Bayesian averaging includes the correctly specified model, the simulation results show that BMA inference generally performs better than any misspecified model alone, especially with respect to the interval coverage performance. In practice, all candidate models can be misspecified and thus one can argue that Bayesian model averaging may not be effective in reducing bias. Intuitively, if some models tend to overestimate and the other models tend to underestimate the parameters of interest, then the Bayesian model averaging will be effective in reducing bias compared to a specific misspecified model. However, we realize that if all models tend to overestimate (or underestimate) the parameters of interest or if the estimates from the incorrect models are far away from the correct model estimates, then the Bayesian model averaging may not be effective in reducing bias. In addition, because the data do not contain information to distinguish between conditional dependence models, one should not expect that the posterior model probabilities to be accurately estimated in practice, casting some doubt on the utility of the BMA estimate in this case.

In this article we consider only homogeneous dependence models which are identifiable from are data setting. Some researchers [43,44] have argued that one can do better in using a non-identifiable model with some informative prior information compared to a less realistic model with strong assumptions that is identifiable. Further research on an expanded model, potentially with heterogeneous dependence structure, may shed more light on the impact of prior misspecification versus model misspecification and the trade-off between an expanded non-identifiable model with less model assumption but more prior assumption and an identifiable model with stronger model assumption but less prior assumption on the estimation of disease prevalence in the case that we discussed.

We assumed that a perfect gold standard (or definitive) test exists, which may limit the usage of the proposed methods, because arguably all diagnostic tests are imperfect and even those with theoretically perfect properties can be rendered imperfect by laboratory or human errors. It may be fruitful for further methodological research to incorporate measurement errors of the third stage gold standard test, e.g., by a sensitivity analysis [45] or multiple imputation [46]. However, this is beyond our present scope.

Another important potential bias in the estimation of disease prevalence is selection bias as to who participates in the screening program. We acknowledge that the estimates from our method can be biased if those who participate in the screening program are not representative of the target population whose prevalence is being estimated. If the information on who tend to participate in the screening program is available, further adjustment for the selection bias can be done by e.g., multiple imputation or inverse probability weighting (i.e., weighting each participant by the inverse of its estimated probability of participating the screening program).

Acknowledgments

Haitao Chu was supported in part by the Lineberger Cancer Center Core Grant CA16086 from the U.S. National Cancer Institute. The authors are grateful to the editor and two anonymous referees for their constructive comments and suggestions which have greatly improved this manuscript.

Reference List

1. Bohning D, Patilea V. A capture-recapture approach for screening using two diagnostic tests with availability of disease status for the test positives only. Journal of the American Statistical Association. 2008; 103:212–21.
2. Vacek PM. The Effect of Conditional Dependence on the Evaluation of Diagnostic-Tests. Biometrics. 1985; 41(4):959–68. [PubMed: 3830260]
3. Torrance-Rynard VL, Walter SD. Effects of dependent errors in the assessment of diagnostic test performance. Statistics in Medicine. 1997; 16(19):2157–75. [PubMed: 9330426]
4. Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. Biometrics. 2001; 57(1):158–67. [PubMed: 11252592]
5. Albert PS, Dodd LE. A cautionary note on the robustness of latent class models for estimating diagnostic error without a gold standard. Biometrics. 2004; 60(2):427–35. [PubMed: 15180668]
6. Qu YS, Tan M, Kutner MH. Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. Biometrics. 1996; 52(3):797–810. [PubMed: 8805757]
7. Qu YS, Hadgu A. A model for evaluating sensitivity and specificity for correlated diagnostic tests in efficacy studies with an imperfect reference test. Journal of the American Statistical Association. 1998; 93(443):920–8.
8. Albert PS, McShane LM, Shih JH. Latent class modeling approaches for assessing diagnostic error without a gold standard: With applications to p53 immunohistochemical assays in bladder tumors. Biometrics. 2001; 57(2):610–9. [PubMed: 11414591]
9. Albert PS. Estimating diagnostic accuracy of multiple binary tests with an imperfect reference standard. Statistics in Medicine. 2009; 28:780–97. [PubMed: 19101935]
10. Albert PS, Dodd LE. On Estimating Diagnostic Accuracy From Studies With Multiple Raters and Partial Gold Standard Evaluation. Journal of the American Statistical Association. 2008; 103(481): 61–73. [PubMed: 19802353]
11. Yang I, Becker MP. Latent variable modeling of diagnostic accuracy. Biometrics. 1997; 53(3): 948–58. [PubMed: 9290225]
12. Espeland MA, Handelman SL. Using Latent Class Models to Characterize and Assess Relative Error in Discrete Measurements. Biometrics. 1989; 45(2):587–99. [PubMed: 2765639]
13. Chu H, Chen S, Louis TA. Random Effects Models in a Meta-Analysis of the Accuracy of Two Diagnostic Tests without a Gold Standard. Journal of the American Statistical Association. 2009; 104:512–23. [PubMed: 19562044]
14. Black MA, Craig BA. Estimating disease prevalence in the absence of a gold standard. Statistics in Medicine. 2002; 21(18):2653–69. [PubMed: 12228883]
15. Hui SL, Walter SD. Estimating the Error Rates of Diagnostic-Tests. Biometrics. 1980; 36(1):167–71. [PubMed: 7370371]
16. Walter SD. Estimation of test sensitivity and specificity when disease confirmation is limited to positive results. Epidemiology. 1999; 10(1):67–72. [PubMed: 9888282]
17. Chu H, Nie L. A few remarks on “A capture-recapture approach for screening using two diagnostic tests with availability of disease status for the test positives only” by Böhning and Patilea. Journal of the American Statistical Association. 2008; 103:1518–9. [PubMed: 20539836]
18. Satten GA, Kupper LL. Inferences About Exposure-Disease Associations Using Probability-Of-Exposure Information. Journal of the American Statistical Association. 1993; 88(421):200–8.
19. Lyles RH. A note on estimating crude odds ratios in case-control studies with differentially misclassified exposure. Biometrics. 2002; 58(4):1034–6. [PubMed: 12495160]
20. Pepe MS, Janes H. Insights into latent class analysis of diagnostic test performance. Biostat. 2007; 8(2):474–84.
21. Ibrahim JG, Lipsitz SR, Chen MH. Missing Covariates in Generalized Linear Models When the Missing Data Mechanism Is Non-Ignorable. Journal of the Royal Statistical Society Series B (Statistical Methodology). 1999; 61(1):173–90.
22. Cheng H, Macaluso M, Waterbor J. Estimation of relative and absolute test accuracy. Epidemiology. 1999; 10(5):566–7. [PubMed: 10468435]

23. Pepe MS, Alonzo TA. Comparing disease screening tests when true disease status is ascertained only for screen positives. Biostat. 2001; 2(3):249–60.
24. Hjort NL, Claeskens G. Frequentist model average estimators. Journal of the American Statistical Association. 2003; 98(464):879–99.
25. Efron, B.; Tibshirani, RJ. An Introduction to the Bootstrap. New York: Chapman & Hall/CRC; 1993.
26. Agresti, A. Categorical data analysis. 2. John Wiley & Sons, Inc; 2002.
27. Gelman, A.; Carlin, JB.; Stern, HS.; Rubin, DB. Bayesian Data Analysis. Chapman & Hall/CRC; 1995.
28. Carlin, BP.; Louis, TA. Bayes and Empirical Bayes Methods for Data Analysis. 2. Chapman & Hall/CRC; 2000.
29. Davidian M, Giltinan DM. Nonlinear models for repeated measurement data: An overview and update. Journal of Agricultural Biological and Environmental Statistics. 2003; 8(4):387–419.
30. Hoeting JA, Madigan D, Raftery AE, Volinsky CT. Bayesian model averaging: A tutorial. Statistical Science. 1999; 14(4):382–401.
31. Gelfand AE, Smith AFM. Sampling-Based Approaches to Calculating Marginal Densities. Journal of the American Statistical Association. 1990; 85(410):398–409.
32. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. Journal of Royal Statistical Society, Series B. 2002; 63(4):583–639.
33. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. Statistical Science. 1992; 138:182–95.
34. Brooks SP, Gelman A. Alternative methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics. 1998; 7:434–55.
35. Natarajan R, McCulloch CE. Gibbs sampling with diffuse proper priors: A valid approach to data-driven inference? Journal of Computational and Graphical Statistics. 1998; 7(3):267–77.
36. Strax P, Venet L, Shapiro S, Gross S. Mammography and Clinical Examination in Mass Screening for Cancer of Breast. Cancer. 1967; 20(12):2184. [PubMed: 6073895]
37. De Sutter P, Coibion M, Vosse M, Hertens D, Huet F, Wesling F, et al. A multicentre study comparing cervicography and cytology in the detection of cervical intraepithelial neoplasia. British Journal of Obstetrics and Gynaecology. 1998; 105(6):613–20. [PubMed: 9647151]
38. Louis TA, Zeger S. Effective Communication of Standard Errors and Confidence Intervals. Biostat. 2009 in press.
39. Pepe, MS. The statistical evaluation of medical tests for classification and prediction. Oxford: Oxford University Press; 2003.
40. Zeger SL, Liang KY. Longitudinal Data-Analysis for Discrete and Continuous Outcomes. Biometrics. 1986; 42(1):121–30. [PubMed: 3719049]
41. Liang KY, Zeger SL. Longitudinal Data-Analysis Using Generalized Linear-Models. Biometrika. 1986; 73(1):13–22.
42. van der Merwe L, Maritz JS. Estimating the conditional false-positive rate for semi-latent data. Epidemiology. 2002; 13(4):424–30. [PubMed: 12094097]
43. Gustafson P. On model expansion, model contraction, identifiability and prior information: Two illustrative scenarios involving mismeasured variables. Statistical Science. 2005; 20(2):111–29.
44. Gustafson P. The utility of prior information and stratification for parameter estimation with two screening tests but no gold standard. Statistics in Medicine. 2005; 24(8):1203–17. [PubMed: 15558709]
45. Chu H, Wang Z, Cole SR, Greenland S. Sensitivity analysis of misclassification: a graphical and a Bayesian approach. Ann Epidemiol. 2006; 16(11):834–41. [PubMed: 16843678]
46. Cole SR, Chu H, Greenland S. Multiple-imputation for measurement-error correction. International Journal of Epidemiology. 2006; 35(4):1074–81. [PubMed: 16709616]

Table 1

Observed frequencies in two screening studies

	Disease Status ($d = 1$)				Disease Status ($d = 2$)				Disease Status ($d = 3$)				Total n
	x^1_{11}	x^1_{10}	x^1_{01}	x^1_{00}	x^2_{11}	x^2_{10}	x^2_{01}	x^1_{00}	x^3_{11}	x^3_{10}	x^3_{01}	x^3_{00}	
Study 1	13	144	95	?	10	24	21	?	—	—	—	—	20,221
Study 2	11	20	81	?	6	29	48	?	14	15	4	?	5,192

Note: Study 1, Health Insurance Plan Screening Study for Breast Cancer in New York. Study 2, a multicenter study for screening cervical cancer.

Table 2

Summary of parameter estimates for conditional dependence using the maximum likelihood method and Bayesian model. The triple notation of ${}_{L}P_{U}$ denotes the point estimate P with 95% Wald-type confidence limits (L, U).

Study	Maximum Likelihood Approach			Bayesian Model		
	$\hat{\alpha}$	$\hat{\theta}$	$\hat{\rho}$	α	θ	ρ
1	${}_{7.65}18.05_{28.45}$	${}_{7.80}14.34_{20.87}$	${}_{0.04}0.09_{0.14}$	${}_{9.42}17.62_{30.7}$	${}_{8.23}13.99_{21.08}$	${}_{0.06}0.10_{0.16}$
2	${}_{4.75}12.94_{21.13}$	${}_{5.02}8.49_{11.96}$	${}_{0.03}0.13_{0.23}$	${}_{5.40}12.36_{21.99}$	${}_{4.62}8.07_{11.71}$	${}_{0.07}0.14_{0.23}$

Note: Study 1, Health Insurance Plan Screening Study for Breast Cancer in New York. Study 2, a multicenter study for screening cervical cancer.

Table 3

Summary of parameter estimates using the maximum likelihood method under the assumption of homogenous dependence. The triple notation of $_{L}P_{U}$ denotes the point estimate P with 95% Wald-type confidence limits (L, U). The estimates of the probabilities of disease classes have been multiplied by 1000 for presentation.

Study	Disease status		Dependent Models		
		Independent Model ($\widehat{P}^{\perp}_{d}$)	the α model ($\widehat{P}^{\alpha}_{d}$)	the θ model ($\widehat{P}^{\theta}_{d}$)	the ρ model ($\widehat{P}^{\rho}_{d}$)
1	No cancer	$_{992}995_{997}$	$_{907}952_{997}$	$_{878}925_{972}$	$_{989}993_{998}$
	cancer	$_{3}5_{8}$	$_{3}48_{93}$	$_{28}75_{122}$	$_{2}7_{11}$
	−2log-likelihood	4056.0	4006.8	4006.8	4006.8
2	Not present	$_{27}51_{75}$	$_{119}389_{658}$	$_{234}424_{614}$	$_{0}115_{308}$
	Low grade	$_{918}942_{966}$	$_{322}594_{867}$	$_{314}515_{716}$	$_{684}878_{1000}$
	High grade	$_{5}7_{10}$	$_{2}17_{32}$	$_{28}61_{94}$	$_{5}8_{11}$
	−2log-likelihood	2756.5	2707.7	2707.7	2707.7

Note: Study 1, Health Insurance Plan Screening Study for Breast Cancer in New York. Study 2, a multicenter study for screening cervical cancer.

Table 4

Summary of posterior estimates using the Bayesian approach under the assumption of homogenous dependence. The triple notation of $_{L}P_{U}$ denotes the posterior median P with 95% equal tailed credible limits (L, U). The posterior estimates of the probabilities of disease classes have been multiplied by 1000 for presentation.

Study	Disease status		Dependent Models			
		Independent Model ($\widehat{P}^{\perp}_{d}$)	the α model ($\widehat{P}^{\alpha}_{d}$)	the θ model ($\widehat{P}^{\theta}_{d}$)	the ρ model ($\widehat{P}^{\rho}_{d}$)	BMA*
1	No cancer	$_{991}995_{997}$	$_{878}953_{982}$	$_{862}926_{962}$	$_{885}992_{996}$	$_{868}955_{995}$
	cancer	$_{3}5_{9}$	$_{18}47_{122}$	$_{38}74_{138}$	$_{4}8_{115}$	$_{5}45_{132}$
2	Not present	$_{34}52_{96}$	$_{153}381_{660}$	$_{215}418_{608}$	$_{54}255_{883}$	$_{68}383_{816}$
	Low grade	$_{897}941_{959}$	$_{321}601_{833}$	$_{329}520_{746}$	$_{108}737_{938}$	$_{175}582_{924}$
	High grade	$_{5}7_{10}$	$_{7}15_{43}$	$_{30}57_{95}$	$_{5}8_{13}$	$_{6}15_{84}$

Note: Study 1, Health Insurance Plan Screening Study for Breast Cancer in New York. Study 2, a multicenter study for screening cervical cancer. BMA = Bayesian model averaging based on the α, θ and ρ dependent models.

Stat Med. Author manuscript; available in PMC 2011 May 20.

Table 5

The means of estimated disease prevalence (true value = 0.2) based on simulation studies with 2000 replicates. The bolded cells represent the correctly chosen model. For the Bayesian models, posterior medians were used as estimates for disease prevalence for a single replicate.

True Model	n	α (θ or ρ)	Maximum Likelihood Methods				Bayesian Models				
			⊥ model	α model	θ model	ρ model	⊥ model	α model	θ model	ρ model	BMA*
⊥ model											
	5000	—	**0.2021**	0.2014	0.1999	0.2023	**0.2026**	0.2003	0.1983	0.2044	0.2019
	25000	—	**0.2002**	0.2001	0.1998	0.2003	**0.2004**	0.2001	0.1996	0.2007	0.2003
α model											
	5000	1.25	0.1815	**0.2013**	0.2120	0.1879	0.1818	**0.2004**	0.2108	0.1897	0.1989
	5000	3.00	0.1256	**0.2001**	0.2656	0.1466	0.1257	**0.1997**	0.2649	0.1475	0.1993
	25000	1.25	0.1801	**0.2004**	0.2124	0.1863	0.1802	**0.2004**	0.2122	0.1867	0.1982
	25000	3.00	0.1253	**0.1999**	0.2664	0.1460	0.1254	**0.2000**	0.2663	0.1461	0.2000
θ model											
	5000	1.25	0.1619	0.1840	**0.1994**	0.1684	0.1620	0.1832	**0.1983**	0.1698	0.1814
	5000	3.00	0.0667	0.0737	**0.1993**	0.0694	0.0667	0.0736	**0.1986**	0.0694	0.0747
	25000	1.25	0.1612	0.1837	**0.2005**	0.1676	0.1613	0.1837	**0.2002**	0.1679	0.1821
	25000	3.00	0.0667	0.0736	**0.2000**	0.0693	0.0667	0.0736	**0.1998**	0.0693	0.0736
ρ model											
	5000	0.20	0.1353	0.3096	0.3712	**0.2042**	0.1357	0.3094	0.3709	**0.2077**	0.3092
	5000	0.60	0.0780	0.5048	0.4613	**0.2130**	0.0789	0.5069	0.4614	**0.2227**	0.4495
	25000	0.20	0.1349	0.3088	0.3716	**0.2008**	0.1349	0.3090	0.3717	**0.2014**	0.3090
	25000	0.60	0.0778	0.5039	0.4608	**0.2017**	0.0778	0.5044	0.4610	**0.2033**	0.4588

*BMA = Bayesian model averaging for the α, θ and ρ dependent models.

Table 6

The 95% confidence/credible interval length of disease prevalence based on simulation studies with 2000 replicates. The bolded cells represent the correctly chosen model.

	n	α (θ or ρ)	Maximum Likelihood Methods				Bayesian Models				
			⊥ model	**α model**	**θ model**	**ρ model**	**⊥ model**	**α model**	**θ model**	**ρ model**	**BMA***
⊥ model											
	5000	—	**0.0706**	0.1062	0.1415	0.0752	**0.0726**	0.1060	0.1394	0.0791	0.1204
	25000	—	**0.0309**	0.0473	0.0633	0.0329	**0.0312**	0.0476	0.0629	0.0332	0.0534
α model											
	5000	1.25	0.0592	**0.1018**	0.1333	0.0686	0.0607	**0.1018**	0.1308	0.0724	0.1149
	5000	3.00	0.0318	**0.0920**	0.1061	0.0500	0.0323	**0.0923**	0.1053	0.0527	0.1736
	25000	1.25	0.0260	**0.0454**	0.0596	0.0300	0.0262	**0.0456**	0.0592	0.0304	0.0604
	25000	3.00	0.0141	**0.0412**	0.0474	0.0220	0.0142	**0.0414**	0.0475	0.0223	0.1454
θ model											
	5000	1.25	0.0485	0.0888	**0.1236**	0.0569	0.0496	0.0890	**0.1210**	0.0597	0.1074
	5000	3.00	0.0141	0.0183	**0.0736**	0.0150	0.0142	0.0184	**0.0729**	0.0150	0.1622
	25000	1.25	0.0215	0.0397	**0.0553**	0.0251	0.0216	0.0400	**0.0550**	0.0253	0.0624
	25000	3.00	0.0063	0.0081	**0.0329**	0.0067	0.0063	0.0082	**0.0326**	0.0067	0.1450
ρ model											
	5000	0.20	0.0369	0.1392	0.1028	**0.1121**	0.0377	0.1383	0.1024	**0.1267**	0.2384
	5000	0.60	0.0163	0.2424	0.0727	**0.2152**	0.0163	0.2380	0.0721	**0.2638**	0.4371
	25000	0.20	0.0163	0.0624	0.0459	**0.0475**	0.0164	0.0624	0.0461	**0.0488**	0.2034
	25000	0.60	0.0073	0.1096	0.0325	**0.0829**	0.0073	0.1089	0.0324	**0.0897**	0.3685

*BMA = Bayesian model average for the α, θ and ρ dependent models.

Stat Med. Author manuscript; available in PMC 2011 May 20.

Table 7

The 95% confidence/credible interval coverage performance of disease prevalence based on simulation studies with 2000 replicates. The bolded cells represent the correctly chosen model.

	n	α (θ or ρ)	Maximum Likelihood Methods				Bayesian Models				
			⊥ model	**α model**	**θ model**	**ρ model**	**⊥ model**	**α model**	**θ model**	**ρ model**	**BMA***
⊥ model											
	5000	—	**0.954**	0.941	0.941	0.956	**0.9525**	0.9470	0.9445	0.9515	0.9680
	25000	—	**0.954**	0.953	0.953	0.952	**0.9560**	0.9590	0.9515	0.9510	0.9740
α model											
	5000	1.25	0.697	**0.945**	0.941	0.828	0.7790	**0.9485**	0.9370	0.9080	0.9725
	5000	3.00	0.000	**0.945**	0.314	0.054	0.0000	**0.9530**	0.2980	0.1350	0.9995
	25000	1.25	0.168	**0.943**	0.867	0.543	0.2065	**0.9445**	0.8570	0.6140	0.9875
	25000	3.00	0.000	**0.952**	0.000	0.000	0.0000	**0.9535**	0.0000	0.0000	1.0000
θ model											
	5000	1.25	0.163	0.825	**0.945**	0.404	0.2405	0.8715	**0.9450**	0.5525	0.9340
	5000	3.00	0.000	0.000	**0.949**	0.000	0.0000	0.0000	**0.9445**	0.0000	0.9070
	25000	1.25	0.000	0.603	**0.946**	0.005	0.0000	0.6525	**0.9475**	0.0090	0.9300
	25000	3.00	0.000	0.000	**0.959**	0.000	0.0000	0.0000	**0.9550**	0.0000	0.9245
ρ model											
	5000	0.20	0.001	0.088	0.000	**0.941**	0.0020	0.0490	0.0000	**0.9400**	0.9120
	5000	0.60	0.000	0.000	0.000	**0.922**	0.0000	0.0000	0.0000	**0.9385**	0.8875
	25000	0.20	0.000	0.000	0.000	**0.952**	0.0000	0.0000	0.0000	**0.9485**	0.9145
	25000	0.60	0.000	0.000	0.000	**0.941**	0.0000	0.0000	0.0000	**0.9515**	0.9250

*BMA = Bayesian model averaging for the α, θ and ρ dependent models.